Digital Libraries from the Perspective of the DELOS Network of Excellence^{*}

Yannis Ioannidis Univ. of Athens, Hellas (Greece) yannis@di.uoa.gr

Abstract

The DELOS Network of Excellence is a European initiative that has been funded by the European Commission under its Fifth and Sixth Framework Programs. It has played a significant role in the formation of an active Digital Libraries research community, the formulation of a vision for the future of the field, and in fostering collaborative research in the direction of this vision. In this paper, we take into account the recommendations of several activities that were initiated by DELOS and outline the DELOS vision for Digital Libraries. Furthermore, we discuss the current research agenda of the Network and the main directions of technical work that it follows.

1. Introduction

Digital Libraries represent the meeting point of a large number of technical areas within the field of informatics, i.e., data management, information retrieval, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction, mass-storage systems, and others. Moreover, Digital Libraries draw upon several other disciplines and fields beyond informatics, such as library sciences, museum sciences, archives, sociology, psychology, etc. Digital Libraries first appeared as a concept in the early 1990s and grew up to become a discipline in its own right through many individual projects that mostly focused on bridging some of the gaps between the constituent fields, understanding what 'digital library functionality' is supposed to be, and integrating solutions from each field into systems that support such functionality.

The results of most of these efforts are systems that manage particular Digital Libraries (collections of content of particular theme(s), to be exact) serving the needs of particular users. Currently, researchers,

0-7803-9228-0/05/\$20.00 ©2005 IEEE

educators, students, and members of other communities continuously search Digital Libraries for information, as part of their daily routines, decision making, or entertainment. Whether they do use the term 'Digital Library' (such as the "Perseus Digital Library" [1], the "ACM Digital Library" [2], or even the "Children's Digital Library" [3]) or not (such as "The Online Books Page" [4]), there are many systems that employ relevant technology. Their functionality has improved the quality of their users' activities' results dramatically, facilitating better information provision and more advanced research. Nevertheless, the current state of affairs is still severely restricted as almost all Digital Libraries developed so far share the following characteristics:

- *Content-centric systems*: Motivation to build Digital Libraries has been based on the need to organize and provide access to particular collections of data and information, more or less ignoring any particular usage requirements.
- *Storage-centric functionality*: Static storage and retrieval of information has been assumed to be the main role of a Digital Library.
- *Isolated and repeated efforts*: Each Digital Library system has been built from scratch, independently of any other development, and reusing nothing that is already available, hence repeating the work associated with several system components that are common.
- *Environment-specific systems*: As a result of the previous characteristic, every Digital Library system has been tailored to the particular needs and characteristics of the target environment with little provision for generalization.
- *Isolated systems*: Even after development, every Digital Library system has mostly operated in isolation, without any interactions with other similar systems. Such fragmentation has limited the opportunities for large-scale analysis and global-scale information availability as well.
- *Application restrictions*: Past Digital Library development has mostly focused on material that is traditionally found in Libraries, and in fact, mostly related to Cultural Heritage.

^{*} Supported by the Information Society Technologies (IST) Program of the European Commission as part of the DELOS Network of Excellence on Digital Libraries (Contract G038-507618).

Hence, despite the undisputed advantages that current Digital Library systems offer compared to the pre-1990s era, the above restrictions limit the role that Digital Libraries can play in Knowledge Societies, which will serve as important educational nuclei in the future.

2. A Vision for Future Digital Libraries

The DELOS Network of Excellence on Digital Libraries (<u>www.delos.info</u>) has initiated a long journey towards filling the gap between current Digital Library practice and the needs of modern information provision. Its goal is to foster the development of technology that will eventually overcome all six of the above restrictions of today's systems and empower everyone to further advance their knowledge, profession, and role in society.

Ultimately, DELOS envisions systems with no logical conceptual, physical, temporal, or personal borders or barriers on information. The DELOS vision is that Digital Libraries will become the universal knowledge repositories and communication conduits for the future, common vehicles by which everyone will access, analyze, evaluate, enhance, and exchange all forms of information [5]. They will be indispensable tools in the daily personal and professional lives of people. They will be accessible at any time and from anywhere, and will offer a friendly, multi-modal, efficient, and effective interaction and exploration environment [6].

Clearly, any effort towards this vision requires significant change in the Digital Library development strategies of the past, with respect to functionality, operational environment, and other aspects. Contrary to current practice, the vision requires that Digital Library systems have the following characteristics:

- *Person-centric efforts*: Humans are at the center of Digital Libraries and all efforts to develop them should be initiated and motivated by needs to provide interesting and/or novel experiences to users. Furthermore, Digital Library systems should synthesize all information that is available about each person in a cohesive whole, so that they may offer personalized treatment to individuals or classes of individuals based on their profiles.
- Communication-centric & collaboration-centric functionality: The main role of Digital Libraries must be to facilitate interaction of scientists, researchers, or the general public on themes that are pertinent to the information stored. Storage of this information and access to it is only a small part of such functionality.
- *Generic technology systems*: For economy of scale, reusability, and extensibility, generic Digital Library Management Systems (DLMSs)

should be developed that capture all common management aspects of Digital Libraries. Supporting any further, environment-specific needs on content manipulation or user interfaces should be developed in a customized fashion on top of DLMSs.

- *Maximum-reuse efforts*: Given the existence of industrial-strength DLMSs, every development effort should depend on them, avoiding much mundane work that is currently necessary, and should only focus on the specialized parts.
- *Globally distributed systems*: Digital Libraries should be managed by widely distributed systems, through which information sources across the world get interconnected to exchange and integrate their contents.
- Universality of information and application: Digital Libraries should be put in the service of ``all'' applications and should comprehensively manage ``all'' forms of content, from data to information to knowledge.

The DELOS Network of Excellence aims at advancing the state of the art in the field so that a first version of the vision may become reality by the end of the decade. It focuses on key instances of the above features and promises to deliver novel technology and robust prototype demonstrators that will be characterized by them. The following section dives a bit deeper into the DELOS workings and outlines the specific research problems that are currently addressed by its members.

3. The DELOS Research Activities

The DELOS research activities are organized around the following main themes that have been identified as critical for realizing the above vision:

Foundational research

• Reference Model for Digital Library Systems Systems-related research

- Digital Library System Architectures
- Information Access to Digital Libraries
- Audio/Visual Digital Libraries
- Semantic Interoperability in Digital Libraries *User-related research*
 - User Interfaces for Digital Libraries
 - Digital Library Visualization
 - Personalization in Digital Libraries

Horizontal issues

- Digital Library Curation and Preservation
- Digital Repositories
- Digital Library Evaluation Methodologies
- Digital Library Evaluation Infrastructures *Applications*
 - E-Health Digital Libraries

Within each of these main themes, there are several problems that are being addressed, which are analyzed in the following subsections.

3.1. Reference Model for Digital Library Systems

Having reached adolescence, the field of Digital Libraries is trying to discover its individual identity. In this direction, DELOS has embarked on an effort to formalize a definition of Digital Library systems, a conceptual framework describing the architectural, functional, operational, and behavioral characteristics of this particular kind of information systems. The outcome of this effort should contribute significantly to everyone's understanding of Digital Library technology and should serve as a yardstick of quality and richness both for the achievements of the field so far and for the generic Digital Library Management Systems that will be developed in the future, since it will specify their expected features and properties. Although initiated by DELOS, this effort will be open for contributions by the entire Digital Libraries community.

3.2. Digital Library System Architectures

A fundamental issue in Digital Library systems is the architecture on top of which they are built to achieve all desirable performance characteristics, i.e., large-scale information distribution, scalability, availability, robustness, reliability, self-organization, adaptability, etc. Recent advances point towards three architectural alternatives: peer-to-peer architectures, grid middleware, and service-oriented architectures. Some combination of them would, of course, be also appropriate.

Peer-to-peer architectures allow for loosely coupled integration of information services and sharing of information such as recommendations and annotations. Grid middleware is appropriate because future Digital Libraries will integrate, in a secure way, information that is broadly distributed. Furthermore, certain services within Digital Libraries are computationally intensive and complex (e.g., extraction of features in multimedia documents to support content-based similarity search or information mining in bio-medical data). Finally, serviceoriented architectures provide mechanisms to describe the semantics and usage of information provision functionality. Moreover, they support mechanisms to combine services into workflow processes for sophisticated search and maintenance.

DELOS studies all three of these architectural directions and evaluates, analytically and experimentally, their impact on Digital Library system architectures.

3.3. Information Access to Digital Libraries

Although not exclusive as in the past, information access will remain a central piece of the functionality of future Digital Libraries. DELOS investigates several research problems that are on the critical path towards development of advanced DLMSs. These may be broadly categorized as problems related to interactions with a single information provider and problems related to integrating information from multiple providers.

Technology has already been developed in other fields for accessing information in individual providers of many complex forms. Indexing, however, remains an uncharted area for several complex and novel data types that arise in Digital Libraries. The kind of measures that are of the necessary complexity to check similarity between two objects require the design of novel data structures that will be based on these measures and will lead to the required data efficiently. The development of such indices is one of the DELOS targets.

With respect to the perennial problem of information integration, the focus of DELOS activities is on query routing in complex distributed Digital Libraries, especially those based on the peer-to-peer architecture. Both schema-based and schema-less information sources have to coordinate and exchange information to respond to user queries, and the main questions are how queries are routed through the system and which criteria are used to choose the best nodes to route queries to.

3.4. Audio/Visual Digital Libraries

Audio/visual information is not found in abundance in traditional libraries but represents a large percentage of the non-traditional objects that are found in Digital Libraries. Offering the appropriate functionality over such information, especially that captured in continuous media, presents unique challenges and DELOS pushes the state of the art on several related issues.

Automatic metadata extraction from audio/visual information is at the forefront of the DELOS effort. Such metadata is essential for efficient search but its manual specification is tedious and extremely time consuming. Any advances in automatic annotation, categorization, segmentation, structure detection, or summary extraction from audio, music, image, or video collections will be very significant.

The second major thrust of work on audio/visual Digital Libraries is content-based retrieval. Using appropriate metadata, whether extracted automatically as anticipated above or not, new algorithms are being developed for query mapping and matching, index-based retrieval, timely delivery, and presentation of information in large media collections. Taking into account user preferences and more general interaction contexts adds further twists to the complexity of these endeavors. Likewise, developing audio/visual user interfaces to arbitrary Digital Libraries shares many of the technical difficulties faced when audio/visual is a characteristic of the contents of the Digital Library, so many of the earlier issues are worked on at the user interface level as well, especially with respect to the natural-language and speech-based interface paradigms.

3.5. Semantic Interoperability in Digital Libraries

Given the widely distributed nature of future Digital Libraries, heterogeneity in the information available is expected to be the norm. Hence, techniques for semantic interoperability will be very important in reconciling differences among existing collections and presenting them as cohesive wholes.

As mentioned earlier, this is an everlasting problem that may never find a complete solution, but DELOS is pushing in certain directions that show promise for marked improvements in the state of the art. Particular emphasis is placed on the investigation of methods for the integration of heterogeneous (upper-level) ontologies and domain-specific knowledge organization systems. By offering formally-grounded mappings between existing ontology standards, transformation of relevant data as well as multiplicity at the conceptual level of user interfaces will become automatically available.

Due to the intimate relationship of Digital Libraries with education, their interoperability with e-Learning applications is part of the DELOS focus as well. The goal is to obtain semantic mappings between learning standards and audio/visual content standards. This will allow the use of semantic descriptions of audio/visual objects to create reusable learning objects.

3.6. User Interfaces for Digital Libraries

Person-centricity of future Digital Libraries puts issues of user interfaces at the center stage of DELOS activities. In the same fashion that the field is trying to identify the core Digital Library system functionality (Section 3.1), it is also lacking a clear understanding of the needs of the users of such systems. DELOS has an on-going effort to collect and systematize user requirements for the different phases of Digital Library usage and identify any non-conventional interaction paradigms that are deemed necessary. The ultimate goal is to build a theoretical framework for the design of new Digital Library user interfaces and use it to implement advanced user-interface layers in future systems.

Currently, the above effort shows a broad and diverse set of activities that are required across various types of Digital Library contents and application environments. The idea of a fixed, one-size-fits-all, interface appears naïve; so, much effort is put on developing technology for task-centered user interfaces, which would self-adapt to the task attempted by the user at any one time.

The importance of collaboration-centric functionality in future systems also drives part of the DELOS effort towards the study of annotations, developing comprehensive annotation models and building compliant systems that will permit users to add their personal opinion, thoughts, and memories to the global Digital Libraries and retrieve those of others.

3.7. Digital Library Visualization

Information visualization usually goes hand-in-hand with user interfaces. The great variety and complexity of information found in Digital Libraries demands for sophisticated visualization techniques as well. Current DELOS work focuses on exploitation of visual clues to make sense out of information presented to the user. Adaptability to small screens as well as interpretation of direct user manipulations of the visualized information as feedback to internal processing mechanisms of the system are two key directions of this work.

3.8. Personalization in Digital Libraries

The requirement for adaptability of system behavior to various user characteristics (e.g., preferences, age, background, etc.) is one more consequence of the personcentricity desired for future DLMSs. Hence, personalization is heavily investigated within DELOS. The primary emphasis is on establishing the appropriate foundations for modeling user preferences and more general user context in Digital Libraries and then using the resulting models to personalize user interactions. Personalization issues, however, are also investigated in conjunctions with other issues, such as personalized query routing in distributed architectures (Section 3.3) or context-aware audio/visual retrieval (Section 3.4).

3.9. Digital Library Curation and Preservation

One of the main distinguishing features of Digital Libraries from other information repositories is that their contents must be valuable at depth of time. This immediately generates the need for long-term preservation of Digital Library material, which is a formidable task, taking into account the speed in which software and hardware evolve and materials deteriorate. Relevant work within DELOS concentrates on integrating preservation functionality into the overall system design, establishing a testbed and evaluation framework for assessing the qualities of preservation techniques and tools, and developing techniques for automating the processes of selection and ingest of materials into a preservation repository.

3.10. Digital Repositories

Digital repositories are, in some sense, a special case of Digital Libraries. DELOS is studying the common ground between the two areas, focusing especially on issues of access policy (mostly open access), operational environment, and the creation and management of Digital Repositories or Libraries. An additional focus would be on the development infrastructure repositories to support research and learning.

3.11. Digital Library Evaluation Methodologies

Not having a precise definition of what a Digital Library is, the community is also lacking a framework for evaluating such systems. The role of evaluation infrastructures, testbeds, and benchmarks is very critical in every field, so DELOS aims to design and develop standard frameworks for comparative evaluation of Digital Library systems. Definition of standard events in a Digital Library environment, identification of appropriate metrics, and establishment of appropriate information repositories are three tasks at the center of this effort. Clearly, this work is closely dependent on the result of the Reference Model effort (Section 3.1).

3.12. Digital Library Evaluation Infrastructures

In the above spirit, DELOS has established and nourished evaluation infrastructures and frameworks on two important aspects of Digital Library functionality: a) Multilingual Information Access: Capturing the world's cultural and scientific heritage, many Digital Libraries contain text in many different languages. Effective access to this information, through interfaces that are equally varied linguistically, is very critical to bringing Digital Libraries to the forefront of Knowledge Society developments. The Cross Language Evaluation Forum (CLEF) is the main international infrastructure and evaluation framework for multilingual information retrieval. Its activities include evaluation campaigns and research into new evaluation methodologies. b) XML Information Access: INEX is an infrastructure and evaluation framework dealing with the performance of systems and services that provide access to XML-based Digital Libraries. Again, INEX activities include evaluation campaigns of actual systems and research into appropriate methodologies and test suites.

3.13. E-Health Digital Libraries

As a very important, non-traditional application of Digital Library technology, DELOS places special emphasis on the area of e-Health, which is in dire need of the future form of Digital Libraries. Large-scale information distribution, heavy computational requirements, privacy, speedy information access, semantic interoperability, and personalization are some of the issues discussed above that become concrete and urgent needs when human health is at stake. DELOS pushes the state of the art in these motivated by the need to integrate, manage, and access virtual electronic health records as well as by the need to integrate and react efficiently to multiple medical information streams, incoming from human-body and environment sensors.

4. Conclusions

The DELOS Network of Excellence is carrying out a broad range of interrelated activities that should lead to the development of the next-generation Digital Library systems, according to the vision outlined in this paper. This represents the collective work of more than 50 DELOS partners, led by a Scientific Board, which is coordinated by Costantino Thanos (CNR-Pisa) and includes Tiziana Catarci (U. Roma "La Sapienza"), Stavros Christodoulakis (T.U. Crete), Alberto del Bimbo (U. Florence), Norbert Fuhr (U. Duisburg), Yannis Ioannidis (U. Athens), Liz Lyon (UKOLN), Seamus Ross (U. Glasgow), and Hans Schek (UMIT). Several DELOS networking activities aim at making the entire Digital Library community part of the overall effort.

References

- [1] <u>www.perseus.tufts.edu</u>
- [2] portal.acm.org/dl.cfm
- [3] www.storyplace.org
- [4] digital.library.upenn.edu/books
- [5] "Digital Libraries at a Crossroads", DELOS FP5 Final Report, July 2003.
- [6] "Digital Libraries: Future Directions for a European Research Program", DELOS Brain-storming Report, San Cassiano, Italy, June 2001.